Science Translational Medicine
AAAS

Supplementary Materials for

Gut microbiome composition may be an indicator of preclinical Alzheimer's disease

Aura L. Ferreiro *et al.*

Corresponding author: Beau M. Ances, bances@wustl.edu; Gautam Dantas, dantas@wustl.edu

Sci. Transl. Med. **15**, eabo2984 (2023)
DOI: 10.1126/scitranslmed.abo2984

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

Supplementary Materials

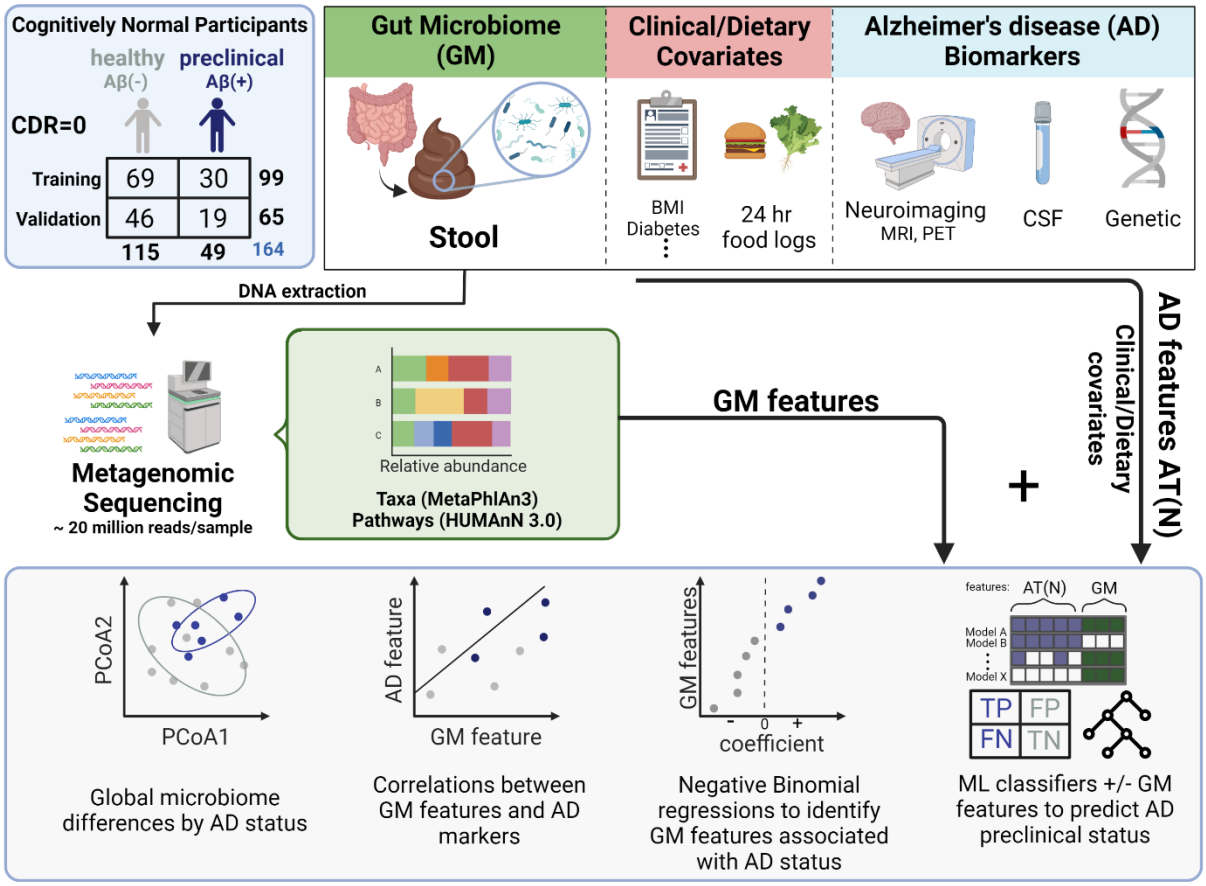
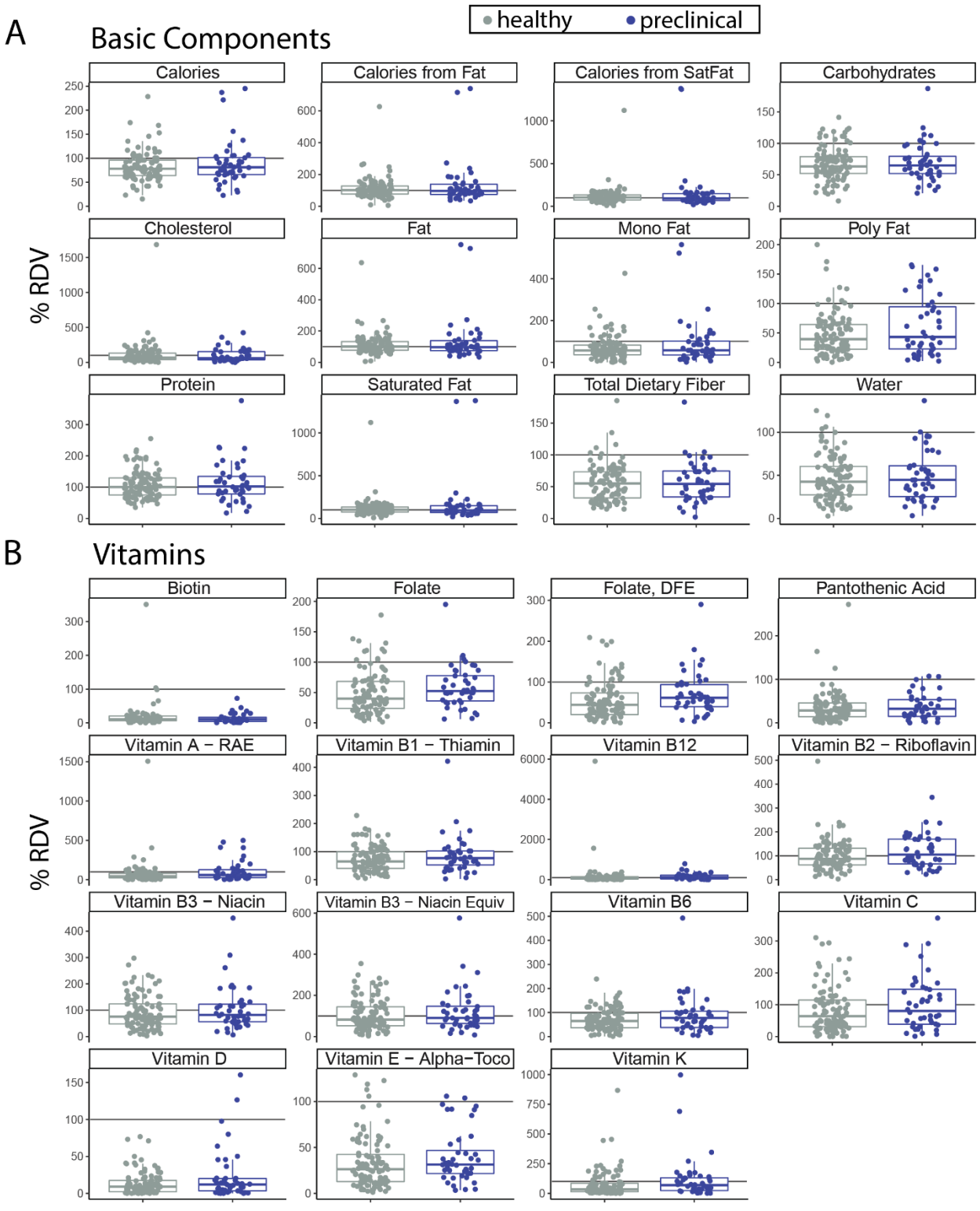


Fig. S1. Schematic of study design and analyses. Stool metagenomes, 24-hour diet logs, clinical covariates, as well as neuroimaging (MRI and PET), CSF, and genetic AD biomarkers from 115 healthy (CDR 0, Aβ marker negative) and 49 preclinical AD (CDR 0, Aβ marker positive) participants were included in this cross-sectional study. For machine learning analyses (Figs 4, S6), participants were randomly split into training (69 healthy, 30 preclinical) or validation (46 healthy, 19 preclinical) cohorts. GM (gut microbiome), AD (Alzheimer's disease), AT(N) (Aβ, tau, and neurodegeneration), ML (machine learning), PCoA (Principal Coordinates Analysis), Aβ (amyloid beta). Schematic created with BioRender.


A
Basic Components
healthy
preclinical
% RDV
Calories
Calories from Fat
Calories from SatFat
Carbohydrates
Cholesterol
Fat
Mono Fat
Poly Fat
Protein
Saturated Fat
Total Dietary Fiber
Water
B
Vitamins
% RDV
Biotin
Folate
Folate, DFE
Pantothenic Acid
Vitamin A − RAE
Vitamin B1 − Thiamin
Vitamin B12
Vitamin B2 − Riboflavin
Vitamin B3 − Niacin
Vitamin B3 − Niacin Equiv
Vitamin B6
Vitamin C
Vitamin D
Vitamin E − Alpha−Toco
Vitamin K

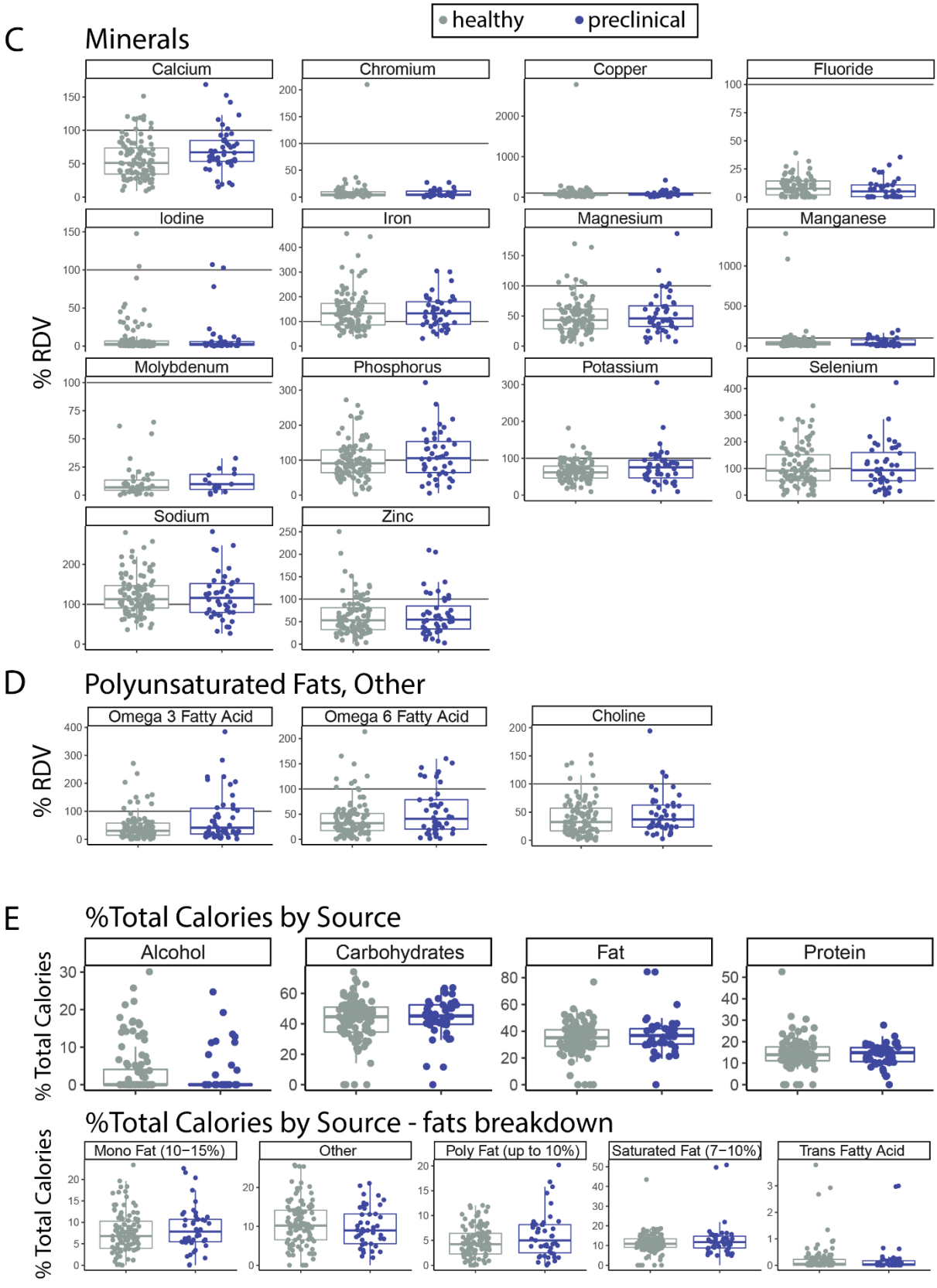

C
Minerals
healthy
preclinical
% RDV
Calcium
Chromium
Copper
Fluoride
Iodine
Iron
Magnesium
Manganese
Molybdenum
Phosphorus
Potassium
Selenium
Sodium
Zinc
D
Polyunsaturated Fats, Other
% RDV
Omega 3 Fatty Acid
Omega 6 Fatty Acid
Choline
E
%Total Calories by Source
% Total Calories
Alcohol
Carbohydrates
Fat
Protein
%Total Calories by Source - fats breakdown
% Total Calories
Mono Fat (10−15%)
Other
Poly Fat (up to 10%)
Saturated Fat (7−10%)
Trans Fatty Acid

Fig. S2. Nutritional profiles from stool sample-matched diet logs do not significantly differ between healthy and preclinical AD groups. Nutritional profiles derived from participant-reported diet logs for the 24 hours preceding stool sample collection are compared between healthy and preclinical AD participants. Percent recommended daily value (% RDV) intakes for healthy and preclinical AD individuals for **(A)** basic dietary components, **(B)** vitamins, **(C)** minerals, and **(D)** polyunsaturated (poly) fat and other nutrient categories. Horizontal gray lines indicate 100% of the RDV for each nutrient. **E**) Percent of total calories sourced from each major nutrient group over the 24-hour period for healthy and preclinical AD individuals. No significant differences were observed for any nutrient category (Student's t-tests, Benjamini-Hochberg adjusted). Nutritional profiles were generated from food logs by a registered and licensed dietitian using the ESHA Food Processor® Nutritional Analysis software.

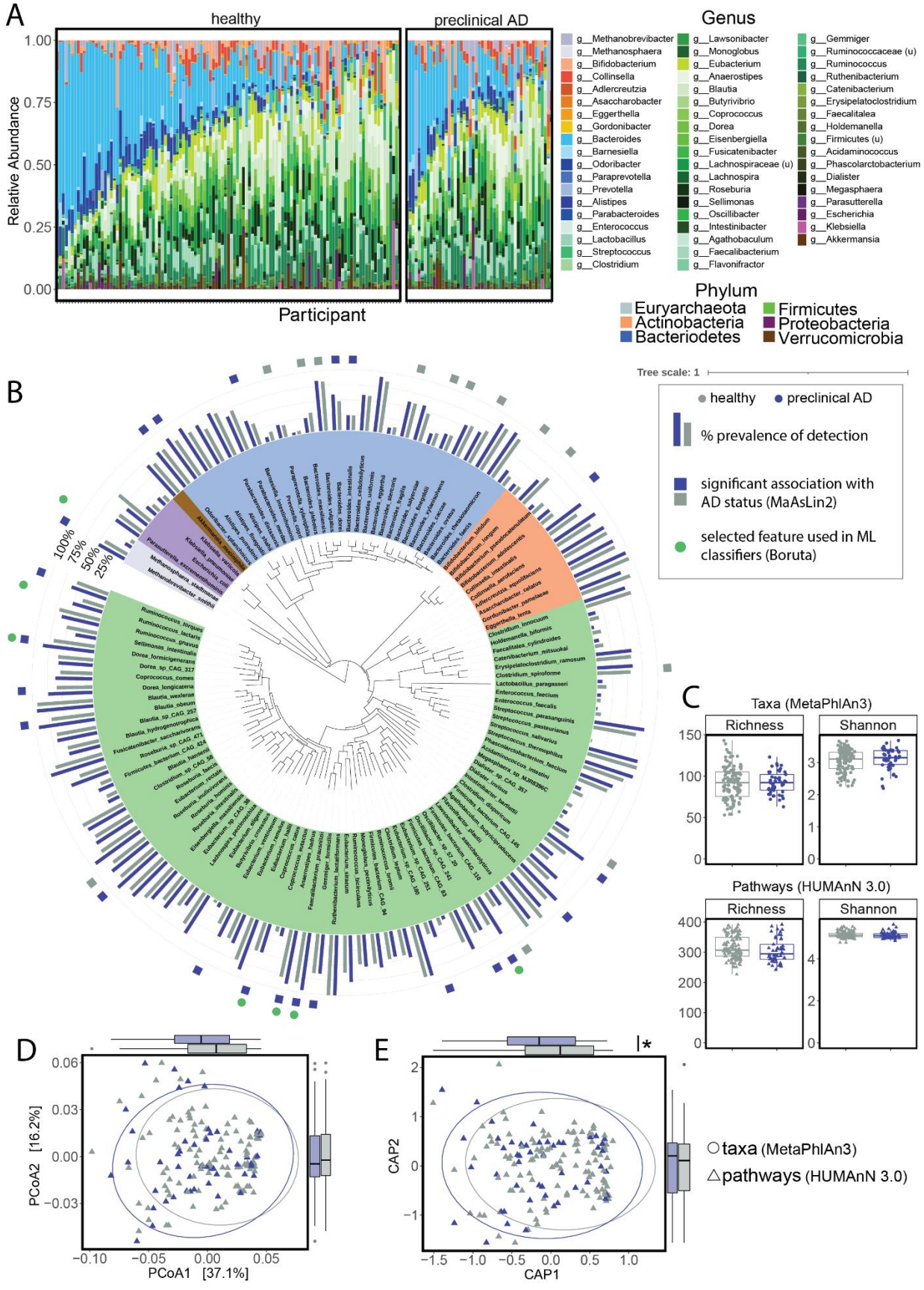

A
healthy
preclinical AD
Relative Abundance
1.00
0.75
0.50
0.25
0.00
Participant
Genus
g__Methanobrevibacter
g__Methanosphaera
g__Bifidobacterium
g__Collinsella
g__Adlercreutzia
g__Asaccharobacter
g__Eggerthella
g__Gordonibacter
g__Bacteroides
g__Barnesiella
g__Odoribacter
g__Paraprevotella
g__Prevotella
g__Alistipes
g__Parabacteroides
g__Enterococcus
g__Lactobacillus
g__Streptococcus
g__Clostridium
g__Lawsonibacter
g__Monoglobus
g__Eubacterium
g__Anaerostipes
g__Blautia
g__Butyrivibrio
g__Coprococcus
g__Dorea
g__Eisenbergiella
g__Fusicatenibacter
g__Lachnospiraceae (u)
g__Lachnospira
g__Roseburia
g__Sellimonas
g__Oscillibacter
g__Intestinibacter
g__Agathobaculum
g__Faecalibacterium
g__Flavonifractor
g__Gemmiger
g__Ruminococcaceae (u)
g__Ruminococcus
g__Ruthenibacterium
g__Catenibacterium
g__Erysipelatoclostridium
g__Faecalitalea
g__Holdemanella
g__Firmicutes (u)
g__Acidaminococcus
g__Phascolarctobacterium
g__Dialister
g__Megasphaera
g__Parasutterella
g__Escherichia
g__Klebsiella
g__Akkermansia
Phylum
Euryarchaeota
Actinobacteria
Bacteriodetes
Firmicutes
Proteobacteria
Verrucomicrobia
B
Tree scale: 1
healthy
preclinical AD
% prevalence of detection
significant association with AD status (MaAsLin2)
selected feature used in ML classifiers (Boruta)
100%
75%
50%
25%
C
Taxa (MetaPhlAn3)
Richness
Shannon
Pathways (HUMAnN 3.0)
Richness
Shannon
D
PCoA2 [16.2%]
PCoA1 [37.1%]
E
CAP2
CAP1
taxa (MetaPhlAn3)
pathways (HUMAnN 3.0)

Fig. S3. Gut microbiome characteristics of the study cohort. A) Stacked taxonomic (MetaPhlAn3) bar plots at the genus level by participant, with color grouping at the phylum level (inset legend). Participants ordered by increasing Firmicutes/Bacteroidetes relative abundance ratio. Lowly abundant taxa exhibiting mean relative abundance ≤ 0.1% across all samples are omitted. 'u': unclassified. **B**) Phylogenetic tree of observed taxa in (A). Taxa labels colored by phylum (see inset legend for panel A). % prevalence: percentage of samples in which each taxon was detected, stratified by AD status. Top taxa significantly associated with preclinical AD or healthy status in negative binomial regression models (MaAsLin2, abs(6oefficient) > 0.15 and prevalence > 15%) are indicated, as are the taxa selected as training features for Random Forest classifiers (Boruta). Tree visualized using iTOL. **C**) Alpha diversity (richness, or number of unique observed taxa, and Shannon Diversity Index) for MetaPhlAn3 taxa, and HUMAnN 3.0 functional pathways. Alpha diversities are not significantly different by preclinical AD status. **D**) PCoA of between-sample binary Bray-Curtis dissimilarities calculated from HUMAnN 3.0 pathway profiles. After accounting for clinical covariates, functional pathway composition is not significantly associated with preclinical AD status (PERMANOVA model summary in Table S4). **E**) Corresponding CAP (canonical analysis of principal coordinates) ordination on binary Bray-Curtis distances calculated from HUMAnN 3.0 pathway profiles, using the same terms as the PERMANOVA in (D). Functional pathway composition is again not significantly associated with preclinical AD status (Table S5). Sample coordinates along the CAP1 axis significantly differ by AD status ($P = 0.036$, Student's t-test). Ellipses represent 95% confidence bounds around group centroids. *$P < 0.05$. Related to Figs. 1, 3, and 4.

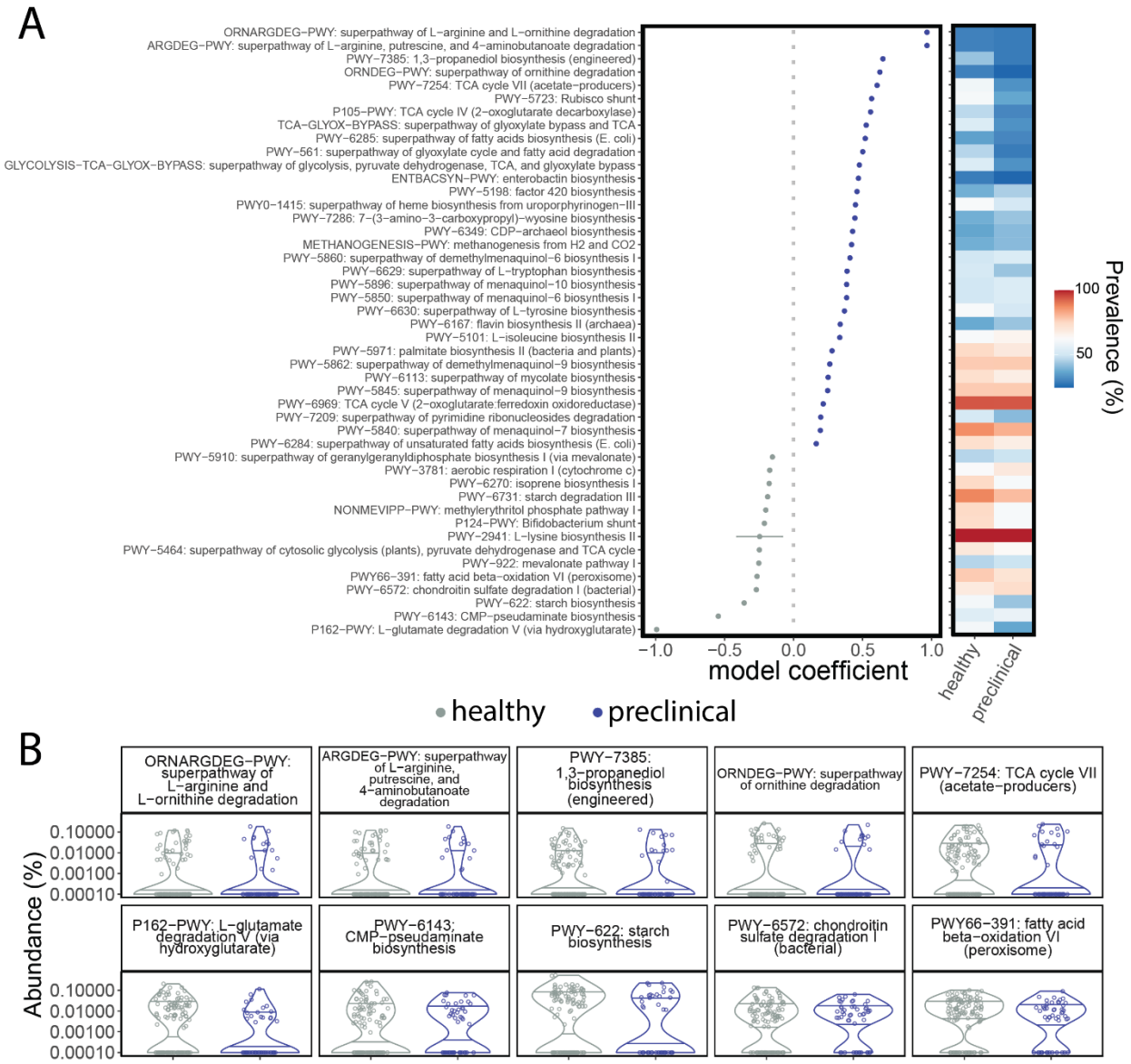


Fig. S4. Fitting negative binomial models to gut microbiome functional pathway data identifies pathways significantly associated with AD preclinical status. **A)** Model coefficients (left panel) and prevalence (right panel) of top-ranking pathways significantly associated with healthy or preclinical AD status. Shown are pathways detected in at least 15% of samples, with Benjamini-Hochberg adjusted p-values of the coefficient < 0.05, and with magnitude of the coefficient > 0.15. Error bars represent the standard error of the coefficient, and may not be visible. Taxa coefficients are from negative binomial regression models (as implemented in MaAslin2)

that additionally include participant age, *APOE* ε4 carrier status, diabetes, body mass index (BMI), hypertension, and time elapsed between PET or lumbar puncture for Aβ quantification and stool collection as predictors. **B)** Relative abundances of the 10 pathways most associated with preclinical AD (top row) or healthy status (bottom row), by their model coefficient. Functional pathway profiles inferred from metagenomic sequence data using HUMAnN3.0. All regression model results available in Supplementary Data File 2. Related to Fig. 3.

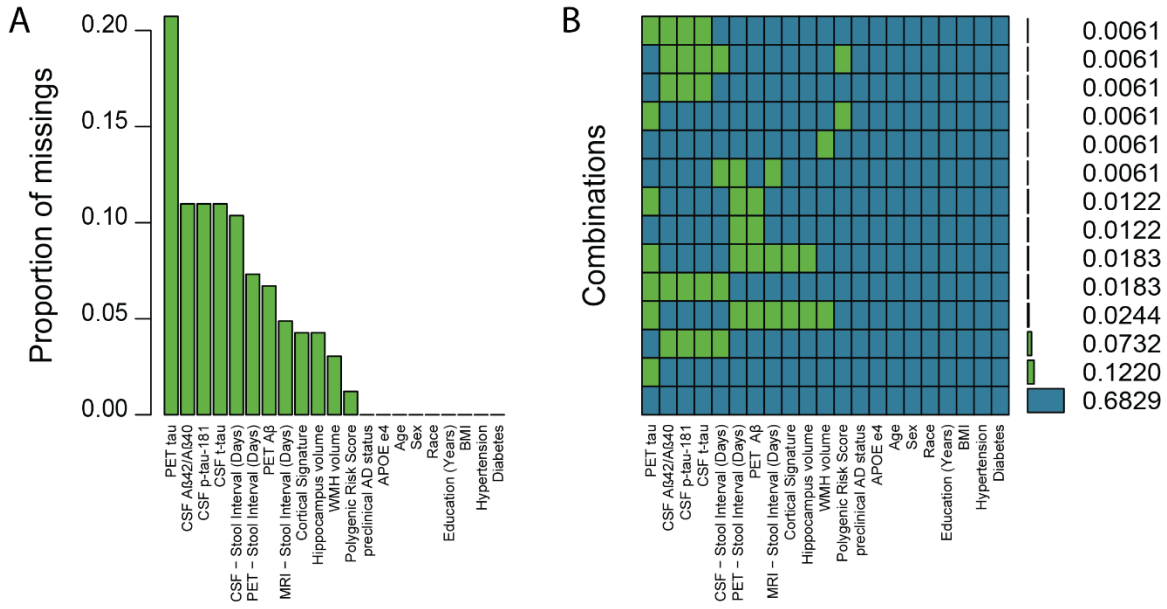


Fig. S5. Summary of missing data imputed for RF analyses. **A)** Proportion of participants with missing data for each feature (biomarker or clinical covariates). **B)** Proportion of missing combinations of features**.** Related to Fig. 4, S6.

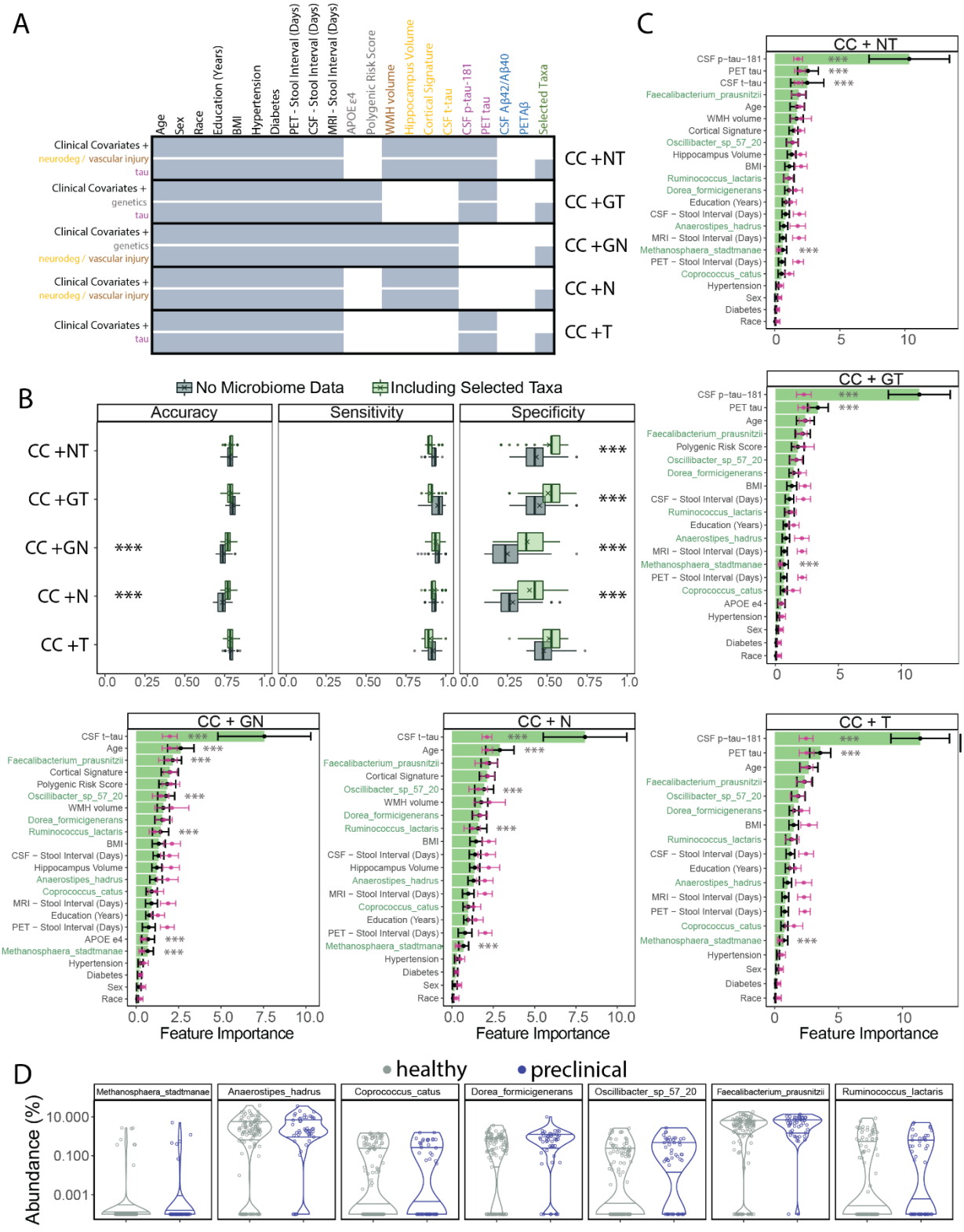

A
Age
Sex
Race
Education (Years)
BMI
Hypertension
Diabetes
PET - Stool Interval (Days)
CSF - Stool Interval (Days)
MRI - Stool Interval (Days)
APOE ε4
Polygenic Risk Score
WMH volume
Hippocampus Volume
Cortical Signature
CSF t-tau
CSF p-tau-181
PET tau
CSF Aβ42/Aβ40
PETAβ
Selected Taxa
Clinical Covariates + neurodeg / vascular injury tau
CC +NT
Clinical Covariates + genetics tau
CC +GT
Clinical Covariates + genetics neurodeg / vascular injury
CC +GN
Clinical Covariates + neurodeg / vascular injury
CC +N
Clinical Covariates + tau
CC +T
B
No Microbiome Data
Including Selected Taxa
Accuracy
Sensitivity
Specificity
CC +NT
CC +GT
CC +GN
CC +N
CC +T
***
0.0 0.25 0.50 0.75 1.0
C
CC + NT
CSF p−tau−181
PET tau
CSF t−tau
Faecalibacterium_prausnitzii
Age
WMH volume
Cortical Signature
Oscillibacter_sp_57_20
Hippocampus Volume
BMI
Ruminococcus_lactaris
Dorea_formicigenerans
Education (Years)
CSF − Stool Interval (Days)
Anaerostipes_hadrus
MRI − Stool Interval (Days)
Methanosphaera_stadtmanae
PET − Stool Interval (Days)
Coprococcus_catus
Hypertension
Sex
Diabetes
Race
0 5 10
CC + GT
Polygenic Risk Score
APOE e4
CC + GN
CC + N
CC + T
Feature Importance
0.0 2.5 5.0 7.5 10.0
0 5 10
D
healthy
preclinical
Abundance (%)
10.000
0.100
0.001
Methanosphaera_stadtmanae
Anaerostipes_hadrus
Coprococcus_catus
Dorea_formicigenerans
Oscillibacter_sp_57_20
Faecalibacterium_prausnitzii
Ruminococcus_lactaris

Fig. S6. Gut microbiome features improve accuracy and specificity, but not sensitivity, of Random Forest (RF) classifiers for AD status that include tau and neurodegeneration biomarkers. A) Summary of features included in each of the RF models reported in (B-C). Feature inclusion is denoted by shaded cells. Compared are models that include or exclude feature-selected gut taxa (bottom and top of each model panel). Feature labels are colored by data/biomarker type (green: gut taxa; blue: amyloid; purple: tau; orange: neurodegeneration; brown: vascular injury; grey: genetic risk factors; black: clinical covariates). All models exclude amyloid biomarkers (PET Aβ and CSF Aβ42/Aβ40). Model shorthand names listed in the right margin. CC: clinical covariates; T: tau; N: neurodegeneration. **B)** Performance metrics for Random Forest models that include or exclude feature-selected gut microbiome taxa (grey: no microbiome features; green: including relative abundances of feature-selected taxa). Box plots summarize performance metrics on the retained validation cohort of models trained on 100 random partitions of the training cohort. Means are denoted by 'X'. $*P < 0.05$, $***P < 0.001$. ANOVAs with Tukey's HSD, Bonferroni adjusted for multiple comparisons at both ANOVA and Tukey honestly significant difference test levels. **C)** Importance of the features included each model, averaged over the 100 training partitions (black), optionally with random class-label shuffling at each iteration to generate null distributions (pink). Error bars represent standard deviation. The 7 taxonomic features are highlighted in green. $***P < 0.001$, Student's t-tests with Benjamini-Hochberg adjustment. **D)** Relative abundances of the 7 feature-selected taxa. Related to Table S9, and Figs. S5, 4.

Table S1. Time elapsed between stool collection and most recent neuroimaging assessments (PET or MRI), lumbar puncture for CSF, or Clinical Dementia Rating (CDR) assessments for healthy and preclinical participants. Bottom row: time elapsed between stool collection and the Aβ marker (from PET or CSF) used to separate participants into preclinical AD and healthy groups. P: Student's t-tests comparing healthy and preclinical groups.

	Healthy mean years (sd)	Preclinical mean years (sd)	P	max (years)	min (days)
PET	2.38 (1.41)	2.30 (1.35)	0.733	7.84	3
CSF	2.58 (1.92)	3.31 (3.38)	0.095	16.6	24
MRI	2.28 (1.47)	2.26 (1.35)	0.934	8.29	3
CDR	0.31 (0.18)	0.34 (0.47)	0.522	3.34	0
Amyloid positivity biomarker (PET or CSF)	2.40 (1.43)	2.35 (1.39)	0.819	10.6	3

Table S2. Permutational analysis of variance (PERMANOVA) test on MetaPhlAn3 taxonomic relative abundances using between-sample UniFrac distances. Terms added sequentially (first to last). Number of permutations: 10,000. Related to Fig 1B**.**

	Df	SumOfSqs	R2	F	P
Age	1	0.201	0.013	2.133	**0.016**
APOE4	1	0.074	0.005	0.790	0.669
Diabetes	1	0.206	0.013	2.196	**0.015**
BMI	1	0.151	0.010	1.602	**0.080**
Hypertension	1	0.037	0.002	0.391	0.986
Interval_Days	1	0.088	0.006	0.932	0.491
amyloid.positive	1	0.174	0.011	1.847	**0.036**
Residual	156	14.670	0.940	NA	NA
Total	163	15.599	1	NA	NA

Table S3. Canonical analysis of principal coordinates (CAP) on the PCoA ordination of MetaPhlAn3 taxonomic relative abundances using between-sample UniFrac distances (Table S2). Terms added sequentially (first to last). Number of permutations: 10,000. Related to Fig 1C**.**

	Df	SumOfSqs	F	P
Age	1	0.217	1.911	**0.018**
APOE4	1	0.095	0.837	0.654
Diabetes	1	0.225	1.975	**0.016**
BMI	1	0.168	1.479	**0.073**
Hypertension	1	0.058	0.505	0.989
Interval_Days	1	0.108	0.945	0.497
amyloid.positive	1	0.191	1.682	**0.040**
Residual	156	17.757	NA	NA

Table S4. Permutational analysis of variance (PERMANOVA) test on HUMAnN 3.0 functional pathway relative abundances using between-sample binary Bray-Curtis distances. Terms added sequentially (first to last). Number of permutations: 10,000. Related to Fig S3D**.**

	Df	SumOfSqs	R2	F	P
Age	1	0.002	0.005	0.816	0.540
APOE4	1	0.008	0.018	2.984	**0.014**
Diabetes	1	0.020	0.042	7.189	**0.000**
BMI	1	0.003	0.007	1.198	0.300
Hypertension	1	0.000	0.000	-0.025	0.998
Interval_Days	1	0.002	0.004	0.672	0.655
amyloid.positive	1	0.003	0.007	1.169	0.298
Residual	156	0.442	0.918	NA	NA
Total	163	0.481	1.000	NA	NA

Table S5. Canonical analysis of principal coordinates (CAP) on the PCoA ordination of HUMAnN 3.0 functional pathway relative abundances using between-sample binary Bray-Curtis distances (Table S4). Terms added sequentially (first to last). Number of permutations: 10,000. Related to Fig S3E**.**

	Df	**SumOfSqs**	**F**	**P**
Age	1	0.004	0.868	0.537
APOE4	1	0.009	2.281	**0.018**
Diabetes	1	0.022	5.204	**0.000**
BMI	1	0.005	1.175	0.262
Hypertension	1	0.002	0.432	0.983
Interval_Days	1	0.003	0.803	0.628
amyloid.positive	1	0.004	1.053	0.353
Residual	156	0.646	NA	NA

Table S6. Model summaries and corresponding analyses of variance for linear regressions of PET Aβ (Centiloid), PET tau (Tauopathy), or Cortical Signature on Axis 1 from the PCoA ordination of MetaPhlAn3 taxonomic abundances (Tax.PCoA1). Dependent variables are regressed on preclinical status ('amyloid.positive'), PCoA1, and their interaction. Reported are the coefficient estimates (s.e.) and [95% CIs]. + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. ANOVAs against null models with preclinical status as the only predictor. Related to Fig 2B, top row panels.

	Model 1: Centiloid~	**Model 2: Tauopathy~**	**Model 3: Cortical Signature~**
(Intercept)	4.721 (1.895)* [0.977, 8.465]	1.172 (0.018)*** [1.137, 1.207]	2.552 (0.011)*** [2.530, 2.573]
Tax.PCoA1	-2.273 (12.307) [-26.591, 22.046]	0.014 (0.117) [-0.217, 0.245]	0.137 (0.071)+ [-0.004, 0.277]
amyloid.positive1	50.975 (3.447)*** [44.163, 57.786]	0.162 (0.033)*** [0.096, 0.227]	-0.033 (0.020) [-0.072, 0.007]
Tax.PCoA1 × amyloid.positive1	-6.143 (24.678) [-54.907, 42.621]	**-0.478 (0.236)*** [-0.944, -0.012]	**-0.255 (0.144)+** [-0.538, 0.029]
Num.Obs.	153	130	157
R2	0.606	0.171	0.048
R2 Adj.	0.598	0.152	0.029
AIC	1344.1	-90.2	-236.0
BIC	1359.3	-75.9	-220.7
Log.Lik.	-667.069	50.117	123.006
ANOVA against null model: (biomarker ~ amyloid.positive)			
Residual DF	149	126	153
RSS	54853	3.521	1.918
DF	2	2	2
Sum of Sq.	69.537	0.144	0.058
F	0.094	2.577	2.296
P	0.910	**0.080**	0.104

Table S7. Model summaries and corresponding analyses of variance for linear regressions of PET Aβ (Centiloid), PET tau (Tauopathy), or Cortical Signature on Axis 2 from the PCoA ordination of MetaPhlAn3 taxonomic abundances (Tax.PCoA2). Dependent variables are regressed on preclinical status ('amyloid.positive'), PCoA2, and their interaction. Reported are the coefficient estimates (s.e.) and [95% CIs]. + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. ANOVAs against null models with preclinical status as the only predictor. Related to Fig 2B, middle row panels.

	Model 1: Centiloid~	**Model 2: Tauopathy~**	**Model 3: Cortical Signature~**
(Intercept)	4.674 (1.850)* [1.019, 8.329]	1.174 (0.017)*** [1.139, 1.208]	2.548 (0.011)*** [2.527, 2.570]
Tax.PCoA2	-11.660 (19.190) [-49.581, 26.260]	0.165 (0.185) [-0.201, 0.530]	-0.112 (0.114) [-0.337, 0.113]
amyloid.positive1	49.669 (3.328)*** [43.094, 56.245]	0.135 (0.032)*** [0.070, 0.199]	-0.033 (0.020) [-0.072, 0.007]
Tax.PCoA2 × amyloid.positive1	**74.039 (34.441)*** [5.983, 142.095]	**0.610 (0.338)+** [-0.058, 1.279]	0.074 (0.206) [-0.333, 0.481]
Num.Obs.	153	130	157
R2	0.619	0.191	0.026
R2 Adj.	0.611	0.172	0.007
AIC	1339.2	-93.3	-232.4
BIC	1354.3	-79.0	-217.1
Log.Lik.	-664.578	51.661	121.206
ANOVA against null model: (biomarker ~ amyloid.positive)			
Residual DF	149	126	153
RSS	53096	3.438	1.951
DF	2	2	2
Sum of Sq.	1826.8	0.227	0.025
F	2.563	4.153	0.962
P	**0.080**	**0.018**	0.385

Table S8. **Model summaries and corresponding analyses of variance for linear regressions of PET Aβ (Centiloid), PET tau (Tauopathy), or Cortical Signature on Axis 1 from the PCoA ordination of HUMAnN 3.0 functional pathway abundances (Fnl.PCoA1).** Dependent variables are regressed on preclinical status ('amyloid.positive'), PCoA1, and their interaction. Reported are the coefficient estimates (s.e.) and [95% CIs]. + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. ANOVAs against null models with preclinical status as the only predictor. Related to Fig 2B, bottom row panels.

	Model 1: Centiloid~	**Model 2: Tauopathy~**	**Model 3: Cortical Signature~**
(Intercept)	4.763 (1.851)* [1.105, 8.420]	1.173 (0.018)*** [1.138, 1.208]	2.550 (0.011)*** [2.528, 2.571]
Fnl.PCoA1	3.104 (59.049) [-113.578, 119.785]	-0.333 (0.552) [-1.424, 0.759]	-0.227 (0.336) [-0.890, 0.437]
amyloid.positive1	48.976 (3.355)*** [42.347, 55.606]	0.138 (0.033)*** [0.073, 0.204]	-0.032 (0.020) [-0.072, 0.008]
Fnl.PCoA1 × amyloid.positive1	**-200.293 (99.685)*** [-397.272, -3.315]	-1.125 (0.944) [-2.993, 0.743]	0.600 (0.588) [-0.562, 1.763]
Num.Obs.	153	130	157
R2	0.621	0.164	0.026
R2 Adj.	0.613	0.144	0.007
AIC	1338.3	-89.1	-232.4
BIC	1353.4	-74.7	-217.2
Log.Lik.	-664.131	49.535	121.222
ANOVA against null model: (biomarker ~ amyloid.positive)			
Residual DF	149	126	153
RSS	52786	3.552	1.962
DF	2	2	2
Sum of Sq.	2136.6	0.112	0.014
F	3.015	1.992	0.526
P	**0.052**	0.141	0.592

Table S9. Summary of statistically significant changes in performance of Random Forest classifiers (Fig. S6) after incorporating gut microbiome features. Mean accuracy, sensitivity, and specificity for RF models trained on subsets of AD biomarkers, with or without gut microbiome features (selected MetaPhlAn3 taxa). Each model was trained on 100 random subsets of the training cohort; shown are the mean performance metrics of those 100 models on the validation cohort. Models are included if they offered significant improvement in at least one performance metric with inclusion of taxonomic features, by ANOVA p-value after Bonferroni adjustment across all ANOVAs (groups: No Microbiome Data, Including Selected Taxa [MetaPhlAn3]). Reported are the corresponding differences of means and 95% CIs. P-values: Tukey's honestly significant difference test after ANOVA for each model, additionally adjusted using the Bonferroni method. Related to Fig. S6. See also Table 2.

		No Microbiome Data		Including Selected Taxa		Including Selected Taxa - No Microbiome Data			
Model	**Metric**	**Mean**	**SD**	**Mean**	**SD**	**Difference**	**CI95 lower**	**CI95 upper**	**P**
Clinical Covariates + neurodeg/vascular + tau	Specificity	0.432	0.090	0.514	0.102	0.082	0.055	0.109	1.05E-07
Clinical Covariates + genetics + tau	Specificity	0.452	0.091	0.505	0.090	0.053	0.027	0.078	8.37E-04
Clinical Covariates + genetics + neurodeg/vascular	Accuracy	0.744	0.026	0.773	0.021	0.030	0.023	0.036	8.60E-13
	Specificity	0.251	0.108	0.374	0.107	0.124	0.094	0.154	1.41E-12
Clinical Covariates + neurodeg/vascular	Accuracy	0.736	0.030	0.771	0.021	0.035	0.028	0.042	8.36E-13
	Specificity	0.284	0.106	0.389	0.100	0.105	0.077	0.134	1.53E-10

Data File S1. Participant metadata (demographics, clinical covariates, and neuroimaging, biofluid, and genetic biomarkers).
Data File S2. MaAsLin2 significant results for taxa and pathways.
Data File S3. Predictive results and performance metrics of Random Forest classifiers for AD preclinical status.
Data File S4. MetaPhlAn3 taxonomic relative abundances, HUMAnN 3.0 pathway relative abundances.